AD_____

Award Number: DAMD17-99-1-9174

TITLE: Computer Aid for the Decision to Biopsy Breast Lesions

PRINCIPAL INVESTIGATOR: Carey Floyd, Ph.D.

CONTRACTING ORGANIZATION: Duke University Medical Center
Durham, North Carolina 27710

REPORT DATE: July 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2000	Annual (1 Jul 99 - 30 Jun 00)

4. TITLE AND SUBTITLE
Computer Aid for the Decision to Biopsy Breast Lesions

5. FUNDING NUMBERS
DAMD17-99-1-9174

6. AUTHOR(S)
Carey Floyd, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Duke University Medical Center
Durham, North Carolina 27710

E-MAIL:
cef@deckard.mc.duke.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This report contains colored photos

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

A case based reasoning (CBR) computer algorithm is under development as a computer aid for the decision to biopsy a breast lesion that is suspected to be cancer. The CBR stores the results of many previous cases along with the results of their biopsies. New cases are evaluated by searching these previous cases and selecting those that are similar to the new cases. A prediction of the results of biopsy is based on the results of the similar known cases. In this report, the algorithm has been implemented on computer and a search has been conducted to optimize the set of image features that are used to compare the new images to the known images. An accuracy of better than 83% is reported and the best performing set of features is described.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

✓ Where copyrighted material is quoted, permission has been obtained to use such material.

✓ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

✓ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Corey E. [illegible] 10 Nov 2000

PI - Signature Date

Table of Contents

Report of the Progress on Grant DAMD17-99-1-9174

For the Period of July 1999 to July 2000

Computer Aid for the Decision to Biopsy Breast Lesions

In the first year of this IDEA award, there was publication activity: one peer-reviewed manuscript was accepted, one reviewed conference proceedings was published, and one presentation was delivered.

Peer-reviewed manuscripts

Floyd C.E., Jr.,Lo J.Y., Tourassi G.D., Breast Biopsy: Case-Based Reasoning Computer-Aid Using Mammography Findings for the Decision to Biopsy, American Journal of Roentgenology (AJR) 175:1-6, 2000.

Reviewed Conference Proceedings

Floyd CE, Jr, Lo JY, Tourassi, GD, "Case-Based Reasoning as a Computer Aid to Diagnosis," Medical Imaging 1999: Image Processing, Hanson KM, Ed., *Proc. SPIE*, 3661:486-489, 1999.

Presentations and abstracts

Floyd CE Jr., Lo JY, Baker JA, Kornguth PJ Multi-Institution Evaluation of Case-Based Reasoning for Breast Cancer Prediction. *Radiolog* 213(P), 334 1999

Narrative

Introduction

A case based reasoning (CBR) system is being developed as a computer aid for the decision to biopsy a lesion for suspected breast cancer. The mammographic findings and patient age are evaluated by the CBR to predict the likelihood of malignancy. This prediction is formed by comparing the case to a knowledge-base of previous cases with

known outcomes. CBR is an intuitive form of computer aided diagnosis since it offers the clinician an accurate, consistent, and interpretable embodiment of diagnostic experience. The focus of this research is to improve the accuracy of breast cancer diagnosis. Breast cancer is usually detected by physical examination or by mammography screening. For women with suspicious findings on their screening mammograms, further diagnostic image studies are usually obtained.

If no definitive diagnosis is obtained from these additional images, the woman and her doctor are faced with two options: biopsy, or short-term follow-up. We propose to improve the accuracy of diagnosis for these women by developing a "Computer Advisor" to predict the likelihood of malignancy from a combination of the findings on the mammograms and the patient history so that this information can be considered when the decision is made.

A long-term goal of our research team is to provide accurate, evidence-based advice to the patient and her health care team at each decision point in this process. This research will establish a decision model to add information after the mammographer has considered all of the available diagnostic evidence and, since cancer was not ruled out by the existing empirical rules, the patient has been referred for biopsy: either excisional (surgery), or needle core. A goal is to demonstrate that the large fraction of benign cases that are referred for biopsy can be reduced and the accuracy of the decision increased by giving the mammographer access to additional information compiled and analyzed by a computer advisor. This additional information can be thought of as an statistical

comparison of this case to a historical archive or knowledge-base of similar cases and their outcomes.

The significance of this problem is demonstrated by the large percentage (66-90%) of breast biopsies that are performed on benign lesions[1]. In the absence of an accurate system for predicting the outcome of biopsy, this large rate of benign biopsies is accepted as a consequence of the effort to correctly identify all malignancies. With this conservative approach, an estimated 2% of cancers that are seen with mammography are incorrectly diagnosed as benign[1].

For a woman with a non-palpable lesion that is visible on her screening mammogram, diagnostic imaging studies including mammography ultrasound and, increasingly, MRI are performed in an effort to rule out or confirm suspicion of breast cancer. When these studies are inconclusive, the patient has the option of biopsy or of waiting and returning later (typically in six-months) for another sequence of images. This option is called short-term follow-up. If the suspicious lesions have remained stable, the region is usually diagnosed as benign. If however, it now appears more malignant, biopsy is typically performed.

Only 10-34% of women who undergo biopsy for non-palpable lesions actually have malignancy[1]. While definitive, unfortunately biopsy can cause complications [2, 3] providing motivation to decrease the number of benign cases referred to biopsy. In addition, about 2% of the referred to short-term follow-up develop cancer at the site of suspicion. The false positive errors (resulting in the benign biopsies) are partially a result of a conservative approach to the decision, driven by the considerable overlap between

those individual mammographic findings seen in both malignant and benign lesions. An accurate decision aid has an opportunity to both increase the low positive predictive value (PPV) of 10-34% by reducing the referral to biopsy of benign cases and to decrease the false negative rate (leading to the referral of malignant cases to follow-up) by correctly referring to biopsy those malignancies that are currently miss-diagnosed.

Our preliminary work suggests that a computer model to predict the outcome of biopsy could form the core of such a decision aid. In this previous work, artificial intelligence techniques were used to help discover non-linear combinations of multiple sources of patient information that successfully predict the outcome of breast biopsy[4-11]. The sources of information include diagnostic findings from mammograms, patient medical history entries, and demographic data (all collectively referred to as findings).

The predictive models proposed for this work is an artificial neural network (ANN) that "learns" to recognize different combinations of findings linked to malignant or benign biopsy outcomes. This technique is data-driven. That is, the combinations of findings and their relationship to benign or malignant outcomes are not specified in the design of the model. No expert rules are built in and the predictive relationships are derived from the data itself. An advantage of such data-driven techniques is that they avoid the bias that can be present in a rule-based model if the rules are based on assumptions that are not optimal. A disadvantage of data-driven techniques is the potential for bias if the database does not accurately represent the population of cases to which the model will ultimately be applied.

A case based reasoning system predicts the likelihood of a malignant biopsy outcome for a new case by considering the question "Of all of the cases seen previously that were similar to this one, what fraction were malignant?" This is a reasonable approach to diagnosis based on clinical experience. There are two advantages to using a computer to address this question. First is consistency. When recalling previous cases, the computer will use the same criteria for deciding which are similar to the current case. Second, the computer has the potential to recall accurately a larger number of cases than any living mammographer could have seen in their career. Third, when implemented within a computerized radiology information system, CBR requires no additional data entry steps for the mammographer and, with one number as an output, provides a consistent, accurate comparison to all previous, similar cases.

The case based reasoning algorithm can be described quite simply. When a new test case is presented for classification, the value of each feature is compared to value of the same feature in the first reference case. If the values of the two features are identical, then the feature is said to match. If the values of every feature is not identical, then a mismatch is counted for each feature that does not match. The sum of the number of features that do not match is recorded as the Hamming distance for each case in the reference data set. The Hamming distance between two cases is defined to be the number of features that do not match exactly. For a given value of the distance cut-off, the matching cases in the reference set are selected as those whose number of features that mismatch is less than or equal to the cut-off. Note that the distance cut-off can take on only integral values. Once all of the matching cases have been identified, the likelihood of malignancy for the new

case is computed as the total number of matching cases that were malignant divided by the total number of matching cases.

Methods

CBR algorithm

CBR predicts an outcome for a new case by examining the outcomes of all similar cases within a knowledge base. In this application, the likelihood of malignancy is predicted as the fraction of all similar cases that were malignant. There are three components to a CBR: a lexicon or coding scheme used to index each case, a knowledge base of cases, and a matching rule to select similar cases. The matching rule uses the lexicon to define similarity between cases.

Lexicon

Mammography cases were indexed using the lexicon of the Breast Reporting System (BI-RADS™) and the patient age. This indexing lexicon has the advantage that it is being used at an increasing number of institutions and thus may allow widespread use of this CBR without requiring any retraining of the mammographers. The BIRADS lexicon consists of categorical and continuous findings. In previous work with artificial neural networks and linear regression analysis, we found that seven findings had the largest contribution to predictive power. These were age, mass margin, mass shape, calcification description, calcification distribution, and associated findings (most significantly the presence of architectural distortion and asymmetric density).

Matching rules

A matching rule is required to select which cases in the knowledge base are similar. Previously we examined the simple rule of requiring all findings of two cases to match exactly. Later [12]this requirement was relaxed by allowing one or more of the findings to differ between two cases. The number of findings that do not match is defined to be the "distance" between the two cases. For categorical data, this distance can have only discrete values. For convenience, the distance between two continuous age findings was discretized by considering the two ages to match if the difference between the two was less than some interval. From previous studies, an interval of three years was chosen. With a distance measure defined, a distance cut off threshold completes the matching rule to determine if two cases are similar. Two cases will be called similar if the number of findings that do not match is less than this threshold. The combination of a set of features and a distance cut off defines a matching rule. In this study, the eight sets of findings described in table 1 and three thresholds (0,1,2) were examined for a total of 24 matching rules.

Knowledge Base

The knowledge base defines the stored experience of the CBR and is formed from archived past cases with known biopsy results. The cases for this project were described for a previous investigation to develop an artificial neural network for the decision to [5]].

Of the women undergoing needle localization for non-palpable breast lesions between January 1991 and December 1995, 500 lesions were randomly selected that went on to open excisional biopsy and pathological diagnosis. These include 206 that were retrospectively read in a previous study[7] and 294 new cases that were prospectively acquired.

Each set of mammograms was acquired using film-screen technique on dedicated mammography equipment. No case was included in the study if either of the reviewing radiologists had prior knowledge of the biopsy results or if the suspicious area was not definitely identified. Of the 500 lesions evaluated there were 232 masses alone, 192 microcalcifications alone, and 29 combinations of masses and associated microcalcifications. The remaining 47 lesions included various combinations of architectural distortion, regions of asymmetric breast density, areas of focal asymmetric density, and areas of asymmetric breast tissue. Patients ranged in age from 24 to 86 years with an average age of 55 years. At biopsy, 326 (65%) of the lesions were found to be benign while 174 (35%) were malignant. This PPV of 35% is greater than reported in prior studies[13, 1, 3, 14], but consistent with our previous data.

All films were read by radiologists whose primary clinical responsibilities are the interpretation of mammograms and the evaluation of breast lesions and who routinely report case findings using the BI-RADS™descriptors. The radiologist was asked to describe each lesion using the BI-RADS™ lexicon by completing a checklist that included all possible BI-RADS™ descriptors. The radiologist was permitted to select only a single descriptor from each category. The findings were recorded during the

routine patient workup before biopsy results were known. The reviewing mammographer was provided with the patient's history and any prior films.

The cases are randomly numbered with no identifying marks that can be traced to the original patients in order to ensure that patient confidentiality is maintained.

Input findings

The input features were selected from ten of the features from the BI-RADS™ lexicon and one finding from the medical history. The ten features initially considered from the BI-RADS™ lexicon were chosen based on our previous work with these data and included mass size, mass margin, mass density, mass shape, calcification description, calcification number, calcification distribution, and special cases/associated findings. The patient's age was included from the history findings. We found that performance strongly depended on which features were included in the matching criteria. No sophisticated feature selection algorithm was used. To reduce the initial number of features, a forward stepwise linear discriminate analysis (LDA) was performed with these eleven potential input features and six were found to contribute at a significance level of 0.05. These selected features were: Age, Mass Margin, Mass Density, Calcification Description, Calcification Distribution, and Associated Findings (including the architectural distortion descriptor). The CBR can be considered a very restricted linear model and so feature exclusion using LDA should include any features useful to CBR.

Output

With a matching rule defined, all cases in the knowledge base that match are selected. The output of the CBR is the fraction of these matching cases that were malignant. A threshold is set on this output to form a binary decision.

Evaluation

The system performance can be evaluated for a given matching distance by sweeping a decision threshold over this likelihood of malignancy from a value of 0 to a value of 1. At each decision value, the true positive fraction and false positive fraction are computed and a receiver operating characteristic curve is drawn. The standard criteria for comparing two diagnostic systems is the area under this ROC curve. For decision to biopsy, this evaluation criteria may be inappropriate since it weights a false positive and a false negative error equally. For breast cancer diagnosis high sensitivity is more important than high specificity. For this reason, we also consider the partial area under the ROC curve over the region between 90 and 100 percent sensitivity. In addition we report the specificity at two values of sensitivity: 100 and 98. While it is customary to use a fitting algorithm to estimate the area under the curves[15], we have found that for these data the standard fitting programs do not accurately represent the data in the regions of high sensitivity. For this reason, the ROC curves were integrated numerically using Newton's method.

To evaluate the contribution of individual findings, the performance of the algorithm was evaluated on a subset of all possible combinations of the six input features. The combinations that were tested are shown in table 1. These combinations represent the

logical choices of grouping for these features. All eight feature combinations were examined and their performance was evaluated for a reasonable range of distance cut off values.

Table 2 Findings included in the matching rules

Findings	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8
Age	X	X	X	X	X	X	X	X
Mass Margin	X	X	X	X	X	X	X	X
Calcification Description	X	X	X	X	X	X	X	X
Mass Density			X	X	X	X		
Calcification Distribution			X	X			X	X
Associated Findings		X		X		X		X

Table 1 The table shows which findings were included in each of the eight matching rules that were tested.

Results

A receiver operating characteristic curve for the CBR performance is shown in fig. 1 below. Note the encouraging behavior at high sensitivity. The sensitivity remains very high as the false positive fraction (FPF) decreases and does not significantly decrease until the FPF has dropped to 0.6 (specificity of 0.4). With a threshold of 0.2, 126 benign biopsies could be avoided at a cost of 2 missed malignancies.

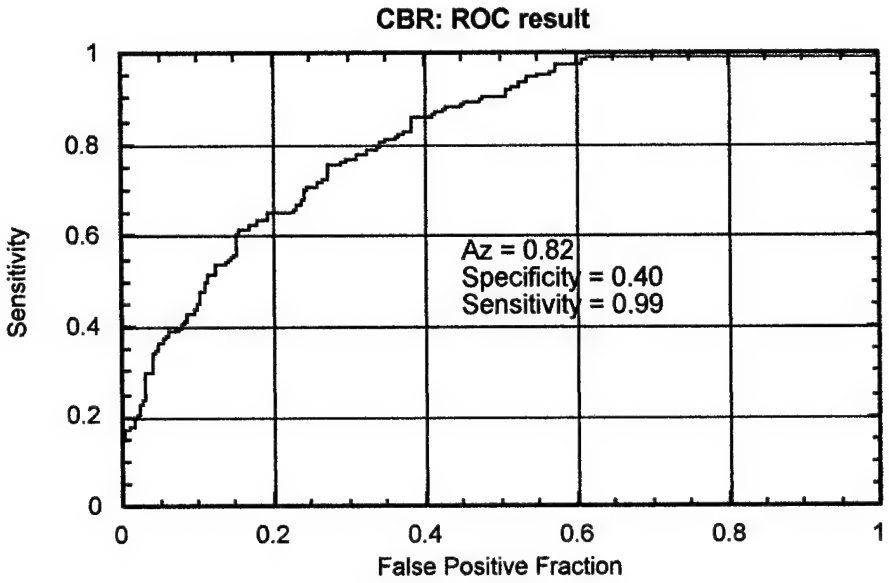

Fig. 1. ROC plot of CBR output values for all benign and malignant cases.

The portion of the ROC curve that is of greatest interest is the region of greatest true-positive fraction (i.e. highest sensitivity) since few radiologists or patients would be willing to under diagnose breast cancer for the sake of high specificity. At sensitivity of 0.98 (relative to all biopsied lesions) the specificity of some of our previous classifiers has been as high as 0.4. Thus, almost 40% the benign biopsies could have been avoided at the cost of missing 2% of the malignancies. The positive predictive value would be increased from 35% to 46%. This study shows that classifiers using the BI-RADS^{TM} lexicon as inputs has the potential to improve the positive predictive value of the recommendation for breast biopsy. The best performance was found for feature set 1. The performance is shown in table 2 for this set.

Best performance: Set 1

Az	Az90	Spec100	Spec98	Spec90
0.82	0.05605	0.25	0.40	0.55

Table 2 Performance for the best set of features.

The inclusion of associated findings was not found to significantly affect the performance and so was eliminated from the feature set. Interestingly, the inclusion of Mass Density and Calcification Distribution were found to degrade the performance.

Discussion

Implementation

This case based reasoning system has been implemented using a relational database running on a workstation running the Windows operating system. In a clinical implementation, the mammographer would examine the mammograms and enter the BI-RADS findings into a radiology information system. These systems are all built with a database as the underlying program. The case based reasoning system would access the findings through the database in the radiology information system and then would compare this case to the stored reference database of previous cases. This comparison could be performed very rapidly and the predicted likelihood of malignancy would be displayed at the data entry workstation for the mammographer to consider.

This technology holds the potential to provide the practicing mammographers with an intelligent "case reference" which would evaluate a clinical case , retrieve relevant archived cases with known outcomes, and summarize the known outcomes for

the similar cases in a form that could help with the decision regarding biopsy. This is an application in which the large storage capacity of the computer can provide the mammographer with access to more cases with their outcomes than any living mammographer would have the opportunity to have seen. If a single mammographer in a busy referral-based medical center had the opportunity to study every case for which a biopsy was performed, they might study 750 cases in a year. If this mammographer was fortunate enough to be so involved over a 40 year career, they might personally be involved with up to 30,000 cases. With a systematic data collection effort, it is reasonable to imagine that the reference data of a CBR system could contain more cases than the most experienced mammographer could see in a lifetime of work. The algorithm was implemented with a user interface using the relational database ACCESS™ (Microsoft Inc, Redmond, Washington). Comparing a new case to the knowledge-base of 1500 cases required 0.08 seconds when running on a 600Mhz Pentium III processor under the Windows98 operating system. No attempt was made to optimize this ACCESS application Evaluating a new case against such a database of 35,000 cases could be performed in fewer than 2 sec using a 600Mhz Pentium III personal computer.

Caveats

There are obvious potential difficulties with the CBR approach. First is the dependence of the technique on uniform use of the BIRADS lexicon by different radiologists. Several studies (ref JAB and Wendy Berg) have described both inter as well as intra observer variability in the assignment of reporting categories when a set of films is read by several mammographers with some repeated readings. In the study by Baker, it was found that

while there were variations in the feature values, the artificial neural network performance at a fixed threshold was fairly stable. The same type of study should be performed with the CBR to evaluate its stability under the expected input variations.

The results reported here only considered eight combinations of BIRADS features from the large possible number of combinations. The fact that the system performance was superior for a small number of features could be interpreted in several ways. First, this study may not have included a sufficient number of cases to fully examine the more subtle contributions of some of the findings. Second, it may be that some of the BIRADS findings do not contribute useful information for this diagnosis. Another reasonable interpretation is that the actual relationship between the multiple features and malignancy is more complex than can be represented by the simple model described in this work. As the number of cases is increased, we will be able to examine these questions with more precision.

When drawing conclusions from this study it is important to recognize that the cases included in both the reference as well as the testing sets are from a specific population. These are cases that were sent to biopsy and neither the distribution of findings nor the relationship between the findings and malignancy should be expected to be representative of all cases examined in diagnostic mammography. The relationships between these different case sets is not known and is the subject of other investigations.

Conclusion

In conclusion, the results from this study indicate that CBR can perform accurately as a predictor of malignancy for mammographically suspicious cases sent to biopsy. This

performance is relatively insensitive to differences between the reference set that is chosen. In addition, for the simple Hamming distance measure, there is little difference in performance between distance measures formed from any of several reasonable subsets of BIRADS findings. After an exhaustive search over the different combinations of eight sets of findings, three distance cutoff thresholds, and two different sets of case data, the performance remained comparable, yet not superior to the performance of an ANN that has been published previously. For the technique to demonstrate improved performance, new reference data and more complex matching criteria will need to be examined.

References

1. Kopans DB. The positive predictive value of mammography. AJR. American Journal of Roentgenology 1992; 158: 521-526.
2. Helvie MA, Ikeda DM, and Adler DD. Localization and needle aspiration of breast lesions: complications in 370 cases. AJR. American Journal of Roentgenology 1991; 157: 711-714.
3. Schwartz GF, Carter DL, Conant EF, Gannon FH, Finkel GC, and Feig SA. Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential. Cancer 1994; 73: 1660-1665.
4. Lo JY, Kim J, Baker JA, and Floyd CE, Jr. *Computer-aided diagnosis of mammography using an artificial neural network: Predicting the invasiveness of breast cancers from image features.* in *SPIE Medical Imaging 1996: Image Processing.* 1996.
5. Lo JY, Baker JA, Kornguth PJ, Iglehart JD, and Floyd CE, Jr. Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. Radiology 1997; 203: 159-163.
6. Lo JY, Baker JA, Kornguth PJ, and Floyd CE, Jr. Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks. Academic Radiology 1999; 6: 10-15.
7. Baker JA, Kornguth PJ, Lo JY, and Floyd CE, Jr. Artificial neural network: improving the quality of breast biopsy recommendations. Radiology 1995; 198: 131-135.

8. Baker JA, Kornguth PJ, Lo JY, Williford ME, and Floyd CE, Jr. Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. Radiology 1995; 196: 817-822.

9. Baker JA, Kornguth PJ, and Floyd CE, Jr. Breast imaging reporting and data system standardized mammography lexicon: observer variability in lesion description. AJR. American Journal of Roentgenology 1996; 166: 773-8.

10. Floyd CE, Jr, Lo JY, Yun AJ, Sullivan DC, and Kornguth PJ. Prediction of breast cancer malignancy using an artificial neural network. Cancer 1994; 74: 2944-2948.

11. Floyd CE, Jr, Lo JY, Tourassi GD, Baker JA, Vitittoe NF, and Vargas-Voracek R. Computer aided diagnosis in thoracic and mammographic radiology. Medical Imaging Technology 1996; 14: 629-634.

12. Floyd CE, Jr, Lo JY, and Tourassi GD. Cased-based reasoning computer algorithm that uses mammographic findings when physicians decide to perform biopsies of the breast. AJR. American Journal of Roentgenology 2000; 175.

13. Ciatto S, Cataliotti L, and Distante V. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. Radiology 1987; 165: 99-102.

14. Hall FM, Storella JM, Silverstone DZ, and Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. Radiology 1988; 167: 353-358.

15. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Investigative Radiology 1989; 24: 234-245.